

1 Supplement1 - Aggregating predictions from experts:
 2 a review of statistical methods, experiments, and applications

3 Thomas McAndrew^{a,*}, Nutch Wattanachit^a, Graham C. Gibson^a, Nicholas G. Reich^a

4 ^a*Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University*
 5 *of Massachusetts at Amherst, Amherst, Massachusetts, USA*

The primary target was categorical	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]
The primary target was continuous	[19, 1, 20, 21, 22, 23, 24, 3, 4, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50]
The primary target was from a time series	[19, 35, 23, 24, 25, 26, 29, 30, 31, 7, 10, 1, 12, 39, 40, 41, 44, 9, 47, 48, 50, 51]
A novel method/model was developed	[20, 35, 21, 23, 24, 3, 28, 29, 30, 31, 7, 32, 1, 36, 11, 40, 13, 42, 43, 44, 14, 15, 46, 49, 18]
The authors implemented a Bayesian technique	[22, 3, 4, 5, 30, 33, 12, 40, 41, 14, 16, 17, 51]
The authors implemented a Frequentist technique	[35, 21, 2, 52, 26, 27, 6, 32, 8, 1, 36, 11, 37, 38, 39, 40, 13, 42, 45, 44, 15, 46, 9, 47, 48, 50]
The model was nonparametric	[2, 52, 23, 24, 4, 25, 26, 28, 10, 34, 14, 9, 49]
The model was parametric	[19, 21, 2, 22, 3, 4, 5, 27, 6, 29, 30, 7, 32, 8, 33, 35, 36, 11, 37, 12, 38, 39, 40, 41, 13, 42, 43, 44, 45, 15, 46, 47, 16, 48, 17, 18, 50, 51]
The model combined point estimates	[35, 2, 52, 24, 25, 27, 6, 31, 7, 32, 8, 10, 34, 11, 12, 40, 42, 43, 44, 45, 14, 46, 9, 47, 16, 48, 17, 18, 50, 51]
The model combined probabilistic distributions	[21, 22, 3, 4, 26, 28, 29, 30, 33, 34, 36, 37, 12, 38, 39, 41, 13, 17, 49]
Experts depended on data that could be updated, revised, or rapidly change	[20, 21, 23, 24, 25, 29, 30, 31, 7, 10, 12, 39, 41, 13, 44, 45, 15, 9, 47, 48, 50]

Table 1: Citations corresponding to Table 2 of the manuscript.

*This work was funded by the National Institute of General Medical Sciences (NIGMS) Grant R35GM119582. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the views of the NIH or the NIGMS. The funders had no role in study design, data collection and analysis, decision to present, or preparation of the presentation.

*Corresponding author

Email addresses: mcandrew@umass.edu (Thomas McAndrew), nwattanachit@umass.edu (Nutch Wattanachit), gcgibson@umass.edu (Graham C. Gibson), nick@schoolph.umass.edu (Nicholas G. Reich)

6 References

- [1] Coo – Roger M Cooke et al. “Out-of-sample validation for structured expert judgment of Asian carp establishment in Lake Erie.” In: *Integrated Environmental Assessment and Management* 10.4 (2014), pp. 522–528. DOI: [10.1002/ieam.1559](https://doi.org/10.1002/ieam.1559).
- [2] Mak – Brenda Mak, Tung Bui, and Robert Blanning. “Aggregating and updating experts’ knowledge: An experimental evaluation of five classification techniques.” In: *Expert Systems with Applications* 10.2 (1996), pp. 233–241. DOI: [10.1016/0957-4174\(95\)00049-6](https://doi.org/10.1016/0957-4174(95)00049-6).
- [3] Bol – Donnacha Bolger and Brett Houlding. “Deriving the probability of a linear opinion pooling method being superior to a set of alternatives.” In: *Reliability Engineering & System Safety* 158 (2017), pp. 41–49.
- [4] Cle – Robert T Clemen and Robert I Winkler. “Aggregating Probability Distributions.” In: *Advances in Decision Analysis: From Foundations to Applications*. 2007. DOI: [10.1017/CB09780511611308.010](https://doi.org/10.1017/CB09780511611308.010).
- [5] Tar – Daniel M Tartakovsky. “Probabilistic risk analysis in subsurface hydrology.” In: *Geophysical Research Letters* 34.5 (2007). DOI: [10.1029/2007GL029245](https://doi.org/10.1029/2007GL029245).
- [6] Han – Anca M Hanea et al. “The Value of Performance Weights and Discussion in Aggregated Expert Judgments.” In: *Risk Analysis* 38.9 (2018), pp. 1781–1794. DOI: [10.1111/risa.12992](https://doi.org/10.1111/risa.12992).
- [7] Sat – Ville A Satopää et al. “Probability aggregation in time-series: Dynamic hierarchical modeling of sparse expert beliefs.” In: *The Annals of Applied Statistics* 8.2 (2014), pp. 1256–1280. DOI: [10.1214/14-AOAS739](https://doi.org/10.1214/14-AOAS739).
- [8] Ada – Ray Adams, Anthony White, and Efe Ceylan. “An Acceptability Predictor for Websites.” In: *International Conference on Universal Access in Human-Computer Interaction*. Springer. 2009, pp. 628–634. DOI: [10.1007/978-3-642-02713-0_66](https://doi.org/10.1007/978-3-642-02713-0_66).
- [9] Gra – Andreas Graefe et al. “Combining forecasts: An application to elections.” In: *International Journal of Forecasting* 30.1 (2014), pp. 43–54. DOI: [10.1016/j.ijforecast.2013.02.005](https://doi.org/10.1016/j.ijforecast.2013.02.005).
- [10] Gra – Andreas Graefe. “Predicting elections: Experts, polls, and fundamentals.” In: *Judgment and Decision Making* 13.4 (2018), p. 334.
- [11] Jan – Dipak Kumar Jana et al. “Interval type-2 fuzzy logic and its application to occupational safety risk performance in industries.” In: *Soft Computing* 23.2 (2019), pp. 557–567. DOI: [10.1007/s00500-017-2860-8](https://doi.org/10.1007/s00500-017-2860-8).
- [12] Abr – Bruce Abramson et al. “Hailfinder: A Bayesian system for forecasting severe weather.” In: *International Journal of Forecasting* 12.1 (1996), pp. 57–71. DOI: [10.1016/0169-2070\(95\)00664-8](https://doi.org/10.1016/0169-2070(95)00664-8).
- [13] Kur – Dorota Kurowicka et al. “Probabilistic Inversion in Priority Setting of Emerging Zoonoses.” In: *Risk Analysis* 30.5 (2010), pp. 715–723. DOI: [10.1111/j.1539-6924.2010.01378.x](https://doi.org/10.1111/j.1539-6924.2010.01378.x).
- [14] Bri – Mario Brito and Gwyn Griffiths. “A Bayesian approach for predicting risk of autonomous underwater vehicle loss during their missions.” In: *Reliability Engineering & System Safety* 146 (2016), pp. 55–67. DOI: [10.1016/j.res.2015.10.004](https://doi.org/10.1016/j.res.2015.10.004).

- 48 [15] Gu – Wei Gu, Thomas L Saaty, and Rozann Whitaker. “Expert System for Ice Hockey
49 Game Prediction: Data Mining with Human Judgment.” In: *International Journal of*
50 *Information Technology & Decision Making* 15.04 (2016), pp. 763–789. DOI: [10.1142/
51 S0219622016400022](https://doi.org/10.1142/S0219622016400022).
- 52 [16] Wan – Liguang Wang and Xueqing Zhang. “Bayesian analytics for estimating risk
53 probability in PPP waste-to-energy projects.” In: *Journal of Management in Engi-*
54 *neering* 34.6 (2018), p. 04018047.
- 55 [17] Bri – Mario Brito et al. “A Behavioral Probabilistic Risk Assessment Framework for
56 Managing Autonomous Underwater Vehicle Deployments.” In: *Journal of Atmospheric*
57 *and Oceanic Technology* 29.11 (2012), pp. 1689–1703. DOI: [10.1175/JTECH-D-12-
58 00005.1](https://doi.org/10.1175/JTECH-D-12-00005.1).
- 59 [18] Bar – Jonathan Baron et al. “Two Reasons to Make Aggregated Probability Forecasts
60 More Extreme.” In: *Decision Analysis* 11.2 (2014), pp. 133–145. DOI: [10.1287/deca.
61 2014.0293](https://doi.org/10.1287/deca.2014.0293).
- 62 [19] Joh2018 – Fred A Johnson et al. “Making do with less: must sparse data preclude
63 informed harvest strategies for European waterbirds?” In: *Ecological Applications* 28.2
64 (2018), pp. 427–441. DOI: [10.1002/eap.1659](https://doi.org/10.1002/eap.1659).
- 65 [20] Li – Wei Li, Yan-ju Liu, and Zhifeng Yang. “Preliminary Strategic Environmental As-
66 sessment of the Great Western Development Strategy: Safeguarding Ecological Secu-
67 rity for a New Western China.” In: *Environmental Management* 49.2 (2012), pp. 483–
68 501. DOI: [10.1007/s00267-011-9794-1](https://doi.org/10.1007/s00267-011-9794-1).
- 69 [21] Kla – Michael Kläs et al. “Support planning and controlling of early quality assurance
70 by combining expert judgment and defect data—a case study.” In: *Empirical Software*
71 *Engineering* 15.4 (2010), pp. 423–454. DOI: [10.1007/s10664-009-9112-1](https://doi.org/10.1007/s10664-009-9112-1).
- 72 [22] Zio – E Zio and GE Apostolakis. “Accounting for expert-to-expert variability: A po-
73 tential source of bias in performance assessments of high-level radioactive waste reposi-
74 tories.” In: *Annals of Nuclear Energy* 24.10 (1997), pp. 751–762. DOI: [10.1016/S0306-
75 4549\(96\)00052-7](https://doi.org/10.1016/S0306-4549(96)00052-7).
- 76 [23] Pet – Dobrila Petrovic, Ying Xie, and Keith Burnham. “Fuzzy decision support system
77 for demand forecasting with a learning mechanism.” In: *Fuzzy Sets and Systems* 157.12
78 (2006), pp. 1713–1725. DOI: doi.org/10.1016/j.fss.2006.03.011.
- 79 [24] Son – Haiyan Song, Bastian Z Gao, and Vera S Lin. “Combining statistical and judg-
80 mental forecasts via a web-based tourism demand forecasting system.” In: *Interna-*
81 *tional Journal of Forecasting* 29.2 (2013), pp. 295–310. DOI: [10.1016/j.ijforecast.
82 2011.12.003](https://doi.org/10.1016/j.ijforecast.2011.12.003).
- 83 [25] Gra – Andreas Graefe et al. “Accuracy of Combined Forecasts for the 2012 Presidential
84 Election: The PollyVote.” In: *PS: Political Science & Politics* 47.2 (2014), pp. 427–
85 431. DOI: <https://doi.org/10.1017/S1049096514000341>.
- 86 [26] Mor – Oswaldo Morales-Nápoles et al. “Characterization of Precipitation through Cop-
87 ulas and Expert Judgement for Risk Assessment of Infrastructure.” In: *ASCE-ASME*
88 *Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*
89 3.4 (2017), p. 04017012. DOI: [10.1061/AJRUA6.0000914](https://doi.org/10.1061/AJRUA6.0000914).
- 90 [27] Bor – Mark E Borsuk. “Predictive Assessment of Fish Health and Fish Kills in the
91 Neuse River Estuary Using Elicited Expert Judgment.” In: *Human and Ecological*
92 *Risk Assessment* 10.2 (2004), pp. 415–434. DOI: [10.1080/10807030490438454](https://doi.org/10.1080/10807030490438454).

- 93 [28] Cai – Mengya Cai et al. “On a Simple and Efficient Approach to Probability Distri-
94 bution Function Aggregation.” In: *IEEE Transactions on Systems, Man, and Cyber-*
95 *netics: Systems* 47.9 (2016), pp. 2444–2453. DOI: [10.1109/TSMC.2016.2531647](https://doi.org/10.1109/TSMC.2016.2531647).
- 96 [29] Kab – Özgür Kabak and Füsün Ülengin. “Aggregating forecasts to obtain fuzzy de-
97 mands.” In: *Computational Intelligence In Decision And Control*. World Scientific,
98 2008, pp. 73–78. DOI: [10.1142/9789812799470_0012](https://doi.org/10.1142/9789812799470_0012).
- 99 [30] Hua – Anqiang Huang et al. “Improving Forecasting Performance by Exploiting Ex-
100 pert Knowledge: Evidence from Guangzhou Port.” In: *International Journal of In-*
101 *formation Technology & Decision Making* 15.02 (2016), pp. 387–401. DOI: [10.1142/
102 S0219622016500085](https://doi.org/10.1142/S0219622016500085).
- 103 [31] Gra – Andreas Graefe. “Accuracy gains of adding vote expectation surveys to a
104 combined forecast of US presidential election outcomes.” In: *Research & Politics* 2.1
105 (2015), pp. 1–5. DOI: [10.1177/2053168015570416](https://doi.org/10.1177/2053168015570416).
- 106 [32] Cab – Enrique Cabello et al. “Combination and Selection of Traffic Safety Expert
107 Judgments for the Prevention of Driving Risks.” In: *Sensors* 12.11 (2012), pp. 14711–
108 14729. DOI: [10.3390/s121114711](https://doi.org/10.3390/s121114711).
- 109 [33] Nev – LC Neves and D Frangopol. “Life-cycle performance of structures: Combining
110 expert judgment and results of inspection.” In: *Proceedings of the 1st International*
111 *Symposium on Life-Cycle Civil Engineering*. 2008, pp. 409–414.
- 112 [34] Fai – Lee Failing, Graham Horn, and Paul Higgins. “Using Expert Judgment and
113 Stakeholder Values to Evaluate Adaptive Management Options.” In: *Ecology and So-*
114 *ciety* 9.1 (2004). URL: <http://www.ecologyandsociety.org/vol9/iss1/art13/>.
- 115 [35] Alh – Juha M Alho. “Estimating the strength of expert judgement: The case of US
116 mortality forecasts.” In: *Journal of Forecasting* 11.2 (1992), pp. 157–167. DOI: [doi.
117 org/10.1002/for.3980110206](https://doi.org/10.1002/for.3980110206).
- 118 [36] Eva – John S Evans et al. “Use of probabilistic expert judgment in uncertainty analysis
119 of carcinogenic potency.” In: *Regulatory Toxicology and Pharmacology* 20.1 (1994),
120 pp. 15–36. DOI: [10.1006/rtph.1994.1034](https://doi.org/10.1006/rtph.1994.1034).
- 121 [37] Hor – Stephen C Hora and Erim Kardeş. “Calibration, sharpness and the weighting
122 of experts in a linear opinion pool.” In: *Annals of Operations Research* 229.1 (2015),
123 pp. 429–450. DOI: [10.1007/s10479-015-1846-0](https://doi.org/10.1007/s10479-015-1846-0).
- 124 [38] Hat – Michel Hathout et al. “Uncertainty and Expert Assessment for Supporting
125 Evaluation of Levees Safety.” In: *3rd European Conference on Flood Risk Management*
126 *(FLOODrisk 2016)*. 2016, p. 6. DOI: [10.1051/e3sconf/20160703019](https://doi.org/10.1051/e3sconf/20160703019).
- 127 [39] Wan – Xiaofeng Wang, Chao Du, and Zuoliang Cao. “Probabilistic inversion tech-
128 niques in quantitative risk assessment for power system load forecasting.” In: *2008*
129 *International Conference on Information and Automation*. IEEE. 2008, pp. 718–723.
130 DOI: [10.1109/ICINFA.2008.4608092](https://doi.org/10.1109/ICINFA.2008.4608092).
- 131 [40] Ren – Zhou Ren-jun and Duan Xian-zhong. “Optimal combined load forecast based on
132 the improved analytic hierarchy process.” In: *Proceedings. International Conference*
133 *on Power System Technology*. Vol. 2. IEEE. 2002, pp. 1096–1100. DOI: [10.1109/
134 ICPST.2002.1047570](https://doi.org/10.1109/ICPST.2002.1047570).

- 135 [41] Man – Chrystal S Mantyka-Pringle et al. “Understanding and predicting the combined
136 effects of climate change and land-use change on freshwater macroinvertebrates and
137 fish.” In: *Journal of Applied Ecology* 51.3 (2014), pp. 572–581. DOI: [10.1111/1365-
138 2664.12236](https://doi.org/10.1111/1365-2664.12236).
- 139 [42] Bal – Peter Baldwin. “Weighting Components of a Composite Score using Naïve Ex-
140 pert Judgments About Their Relative Importance.” In: *Applied Psychological Mea-
141 surement* 39.7 (2015), pp. 539–550. DOI: [10.1177/0146621615584703](https://doi.org/10.1177/0146621615584703).
- 142 [43] Zio – E Zio. “On the use of the analytic hierarchy process in the aggregation of expert
143 judgments.” In: *Reliability Engineering & System Safety* 53.2 (1996), pp. 127–138.
144 DOI: [10.1016/0951-8320\(96\)00060-9](https://doi.org/10.1016/0951-8320(96)00060-9).
- 145 [44] Bae – Philippe Baecke, Shari De Baets, and Karlien Vanderheyden. “Investigating
146 the added value of integrating human judgement into statistical demand forecasting
147 systems.” In: *International Journal of Production Economics* 191 (2017), pp. 85–96.
148 DOI: [10.1016/j.ijpe.2017.05.016](https://doi.org/10.1016/j.ijpe.2017.05.016).
- 149 [45] Sei – Matthias Seifert and Allègre L Hadida. “On the relative importance of linear
150 model and human judge(s) in combined forecasting.” In: *Organizational Behavior and
151 Human Decision Processes* 120.1 (2013), pp. 24–36. DOI: [10.1016/j.obhdp.2012.
152 08.003](https://doi.org/10.1016/j.obhdp.2012.08.003).
- 153 [46] Mu – L Mu and W Xianming. “Multi-hierarchical durability assessment of existing
154 reinforced-concrete structures.” In: *Proceedings of the 8th International Conference
155 on Durability of Building Materials and Components*. 1999, pp. 49–69. URL: [https:
156 //www.irbnet.de/daten/iconda/CIB1808.pdf](https://www.irbnet.de/daten/iconda/CIB1808.pdf).
- 157 [47] Alv – Jorge Alvarado-Valencia et al. “Expertise, credibility of system forecasts and
158 integration methods in judgmental demand forecasting.” In: *International Journal of
159 Forecasting* 33.1 (2017), pp. 298–313. DOI: [10.1016/j.ijforecast.2015.12.010](https://doi.org/10.1016/j.ijforecast.2015.12.010).
- 160 [48] Shi – Juneseuk Shin, Byoung-Youl Coh, and Changyong Lee. “Robust future-oriented
161 technology portfolios: Black-Litterman approach.” In: *R&D Management* 43.5 (2013),
162 pp. 409–419. DOI: [10.1111/radm.12022](https://doi.org/10.1111/radm.12022).
- 163 [49] Hor – Stephen C Hora et al. “Median Aggregation of Distribution Functions.” In:
164 *Decision Analysis* 10.4 (2013), pp. 279–291. DOI: [10.1287/deca.2013.0282](https://doi.org/10.1287/deca.2013.0282).
- 165 [50] Fra – Philip Hans Franses. “Averaging Model Forecasts and Expert Forecasts: Why
166 Does It Work?” In: *Interfaces* 41.2 (2011), pp. 177–181. DOI: [10.1287/inte.1100.
167 0554](https://doi.org/10.1287/inte.1100.0554).
- 168 [51] Cra – Peter S Craig et al. “Bayesian Forecasting for Complex Systems Using Com-
169 puter Simulators.” In: *Journal of the American Statistical Association* 96.454 (2001),
170 pp. 717–729. DOI: [10.1198/016214501753168370](https://doi.org/10.1198/016214501753168370).
- 171 [52] Hur – WJ Hurley and DU Lior. “Combining expert judgment: On the performance
172 of trimmed mean vote aggregation procedures in the presence of strategic voting.” In:
173 *European Journal of Operational Research* 140.1 (2002), pp. 142–147. DOI: [10.1016/
174 S0377-2217\(01\)00226-0](https://doi.org/10.1016/S0377-2217(01)00226-0).
- 175 [53] Mor – M Granger Morgan. “Use (and abuse) of expert elicitation in support of decision
176 making for public policy.” In: *Proceedings of the National academy of Sciences* 111.20
177 (2014), pp. 7176–7184. DOI: [10.1073/pnas.1319946111](https://doi.org/10.1073/pnas.1319946111).